DATA EXPLORATION – EUROPEAN SOCCER DATABASE

---

1. How many days have passed from the oldest **Match** to the most recent one (dataset time interval)?

ANSWER: 2868 giorni

QUERY:

```
SELECT
        MAX(date) AS Newest_date,
        MIN(date) AS Oldest_date,
        TIMESTAMP_DIFF(MAX(date),MIN(date), DAY) AS Days_Dataset_inteval
FROM `progetto-di-prova-365418.Final_Exercise.Match`
```

---

2. Produce a table which, for each Season and **League** Name, shows some statistics (min, avg, mid-range, max, sum) about the <u>home</u> goals scored. Which combination of Season-League has the highest number of home goals?

ANSWER: England Premier League - Season 2009/2010 con 645 home goals.

QUERY:

```
SELECT match.season,
     leagues.name AS league_name,
     MIN(match.home_team_goal) AS min_homegoals,
     MAX(match.home_team_goal) AS max_homegoals,
     ROUND(AVG(match.home_team_goal),2) AS avg_homegoals,
     (MIN(match.home_team_goal) + MAX(match.home_team_goal))/2 AS
midrange_homegoals,
     SUM(match.home_team_goal) AS sum_homegoals,
FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
on match.league_id = leagues.id
```

```
GROUP BY match.season, league_name
ORDER BY sum_homegoals DESC
LIMIT 1;
```

---

3. Find out how many unique seasons there are in the **Match** table.
Then write a query that shows, for each Season, the number of matches
played by each League. Do you notice anything out of the ordinary?

ANSWER: Il numero di stagioni presenti nella tabella "match" è 8. Si nota che nella
stagione 2013/2014 nella Belgium Jupiler League sono state giocate soltanto 12
partite.

QUERY:

```
SELECT COUNT(DISTINCT season) AS distinct_season
FROM `progetto-di-prova-365418.Final_Exercise.Match`;

SELECT
      match.season,
      leagues.name AS league_name,
      COUNT(match.id) AS number_matches,
FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
on match.league_id = leagues.id
GROUP BY match.season, league_name
ORDER BY number_matches;
```

---

4. Using Players as the starting point, create a new table (PlayerBMI) and add:
   a. a new variable that represents the players' weight in kg (divide the
      mass value by 2.205) and call it kg_weight;
   b. a variable that represents the height in metres (divide the cm value
      by 100) and call it m_height;
   c. a variable that shows the body mass index (BMI) of the player;
   d. Filter the table to show only the players with an optimal BMI (from
      18.5 to 24.9).

How many rows does this table have? How many players do not have an optimal BMI?

ANSWER: I giocatori con un ottimo BMI sono 10197. I giocatori che non hanno un ottimo BMI sono 863.

QUERY:

```sql
CREATE TABLE `progetto-di-prova-365418.Final_Exercise.PlayerBMI` AS
SELECT *,
    weight/2.205 AS kg_weight,
    height/100 AS m_height,
    (weight/2.205) / ((height/100)*(height/100)) AS BMI
FROM `progetto-di-prova-365418.Final_Exercise.Player`;

SELECT COUNT(*) AS N_player_with_optimal_BMI,
FROM `progetto-di-prova-365418.Final_Exercise.PlayerBMI`
WHERE BMI >= 18.5 AND BMI <= 24.9;

SELECT COUNT(*) AS N_player_with_not_optimal_BMI,
FROM `progetto-di-prova-365418.Final_Exercise.PlayerBMI`
WHERE BMI < 18.5 OR BMI > 24.9;
```

---

5. Which **Team** has scored the highest <u>total</u> number of goals (home + away) during the most recent available season? How many goals has it scored?

ANSWER: La squadra che ha fatto più goal nella stagione più recente (2015/16) è il Barcellona con 112 goal.

QUERY:

```sql
SELECT MAX(DISTINCT season) AS season_most_recent
FROM `progetto-di-prova-365418.Final_Exercise.Match`;

SELECT league_name, team_name, SUM(team_goal) AS tot_team_goal
FROM(
    SELECT * FROM
        (SELECT
```

```
              leagues.name AS league_name,
              team_home.team_long_name AS team_name,
              SUM(match.home_team_goal) AS team_goal,
     FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
     LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
     on match.league_id = leagues.id
     LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team_home
     on match.home_team_api_id = team_home.team_api_id
     WHERE season = "2015/2016"
     GROUP BY league_name, team_name
     ) AS table_home_goals
   UNION ALL
   SELECT * FROM
     (SELECT
              leagues.name AS league_name,
              team_away.team_long_name AS team_name,
              SUM(match.away_team_goal) AS team_goal,
     FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
     LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
     on match.league_id = leagues.id
     LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team_away
     on match.away_team_api_id = team_away.team_api_id
     WHERE season = "2015/2016"
     GROUP BY league_name, team_name
     ) AS table_away_goals
   )
GROUP BY league_name, team_name
ORDER BY tot_team_goal DESC
LIMIT 1;
```

---

6. Create a query that, for each season, shows the name of the team that
   <u>ranks</u> first in terms of <u>total</u> goals scored.
   Which team was the one that ranked first in most of the seasons?

ANSWER: La squadra che risulta più volte prima per goals fatti nelle stagioni disputate (tra quelle disponibili nel Dataset) è il "Real Madrid CF" (4 volte su 8 stagioni), seguita da "FC Barcellona" (3 volte su 8 stagioni) e da "Ajax" (1 volta su 8 stagioni).

QUERY:

```sql
SELECT team_name, count(team_name) as num_seasons_as_1stForGoalScored,
FROM(
  SELECT *,
        RANK() over(partition by season order by tot_team_goal DESC) AS
Frist_team_per_goal_each_season
  FROM(
      SELECT season, league_name, team_name, SUM(team_goal) AS tot_team_goal,
      FROM(
          SELECT * FROM
            (SELECT
                    match.season,
                    leagues.name AS league_name,
                    team_home.team_long_name AS team_name,
                    SUM(match.home_team_goal) AS team_goal,
            FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
            LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
            on match.league_id = leagues.id
            LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team_home
            on match.home_team_api_id = team_home.team_api_id
            GROUP BY match.season, league_name, team_name
            ) AS table_home_goals
          UNION ALL
          SELECT * FROM
            (SELECT
                    match.season,
                    leagues.name AS league_name,
                    team_away.team_long_name AS team_name,
                    SUM(match.away_team_goal) AS team_goal,
            FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
            LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
            on match.league_id = leagues.id
            LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team_away
            on match.away_team_api_id = team_away.team_api_id
            GROUP BY match.season, league_name, team_name
            ) AS table_away_goals
          )
      GROUP BY season, league_name, team_name
      )
    )
WHERE Frist_team_per_goal_each_season = 1
GROUP BY team_name;
```

7. From the query above create a new table (TopScorer) containing the top 10 teams in terms of total goals scored.

Then write a query that shows all the possible "pair combinations" between those 10 teams. How many "pair combinations" did it generate?

ANSWER: Le prime 10 squadre per goal che compongono la tabella "TopScorer" sono: FC Barcelona, Real Madrid CF, Celtic, FC Bayern Munich, PSV, Ajax, FC Basel, Manchester City, Chelsea, Manchester United.

Considerano sia i match di andata che i match di ritorno in cui si affrontano le squadre "TopScorer" in ciascun campionato, ci sono un totale di 80 match (10 per ogni stagione tra quelle disponibili nel dataset).

QUERY:

```sql
CREATE TABLE `progetto-di-prova-365418.Final_Exercise.TopScorer` AS
SELECT team_name, SUM(team_goal) AS tot_team_goal,
FROM(
    SELECT * FROM
      (SELECT
            match.season,
            leagues.name AS league_name,
            team_home.team_long_name AS team_name,
            SUM(match.home_team_goal) AS team_goal,
      FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
      LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
      on match.league_id = leagues.id
      LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team_home
      on match.home_team_api_id = team_home.team_api_id
      GROUP BY match.season, league_name, team_name
      ) AS table_home_goals
    UNION ALL
    SELECT * FROM
      (SELECT
            match.season,
            leagues.name AS league_name,
            team_away.team_long_name AS team_name,
            SUM(match.away_team_goal) AS team_goal,
      FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
```

```sql
        LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
        on match.league_id = leagues.id
        LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team_away
        on match.away_team_api_id = team_away.team_api_id
        GROUP BY match.season, league_name, team_name
        ) AS table_away_goals
    )
GROUP BY team_name
ORDER BY tot_team_goal DESC
LIMIT 10;



SELECT season, COUNT(match_id) AS NumTopMatches
FROM (
        SELECT  id as match_id,
                season,
                home_team_api_id,
                away_team_api_id,
                team_long_name_home,
                team_long_name_away,
                CASE WHEN team_long_name_home IN (SELECT team_name FROM
`progetto-di-prova-365418.Final_Exercise.TopScorer`) THEN "TopScorer"
                    ELSE "NormalScorer"
                    END AS ScorerType_home,
                CASE WHEN team_long_name_away IN (SELECT team_name FROM
`progetto-di-prova-365418.Final_Exercise.TopScorer`) THEN "TopScorer"
                    ELSE "NormalScorer"
                    END AS ScorerType_away
        FROM (
                SELECT match.*, team.team_long_name as team_long_name_away
                FROM(
                        SELECT match.*, team.team_long_name as team_long_name_home
                        FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
                        LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS
team
                            on  match.home_team_api_id = team.team_api_id
                        ) as match
                LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team
                    on  match.away_team_api_id = team.team_api_id
            )
        )
WHERE ScorerType_home = "TopScorer" AND ScorerType_away = "TopScorer"
GROUP BY season
```